Artificial Intelligence

AlexNet

Pham Viet Cuong
Dept. Control Engineering & Automation, FEEE
Ho Chi Minh City University of Technology

AlexNet

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

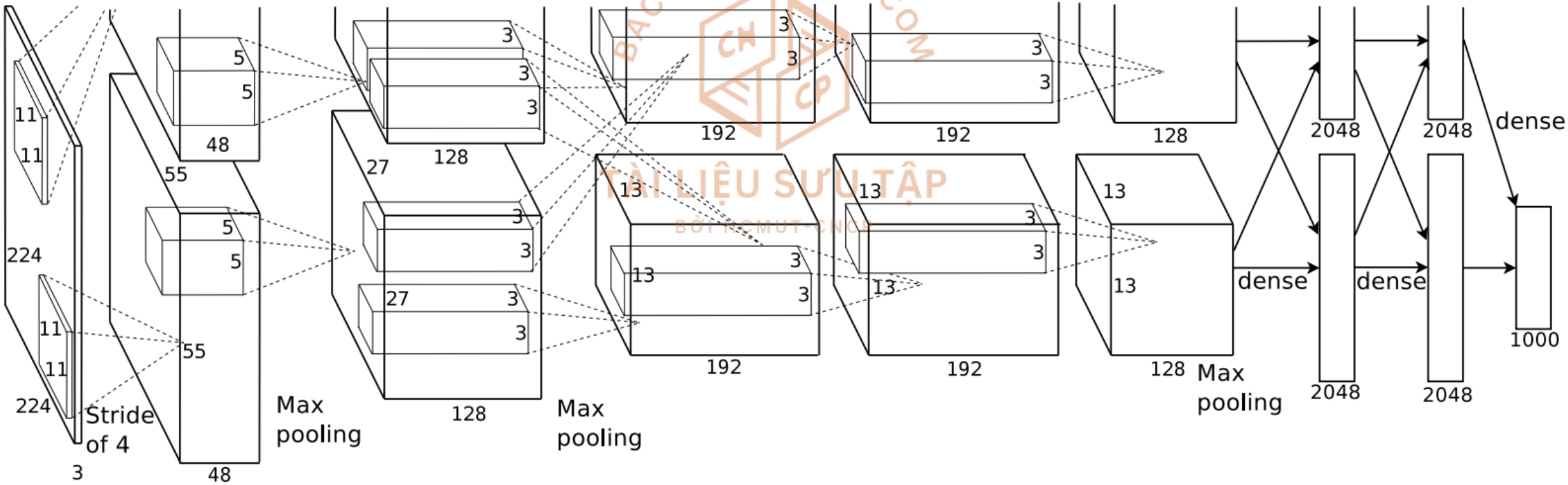
Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

AlexNet

- ✓ Winner of ImageNet LSVRC-2012
- ✓ Image classification problem
  - 1000 classes
  - 650K neurals
  - 62M parameters
  - Training: 1.2M images
  - Validation: 50K images
  - Test: 150K images

Architecture

- ✓ 5 convolutional layers
- ✓ 3 fully-connected layers
- ✓ Softmax

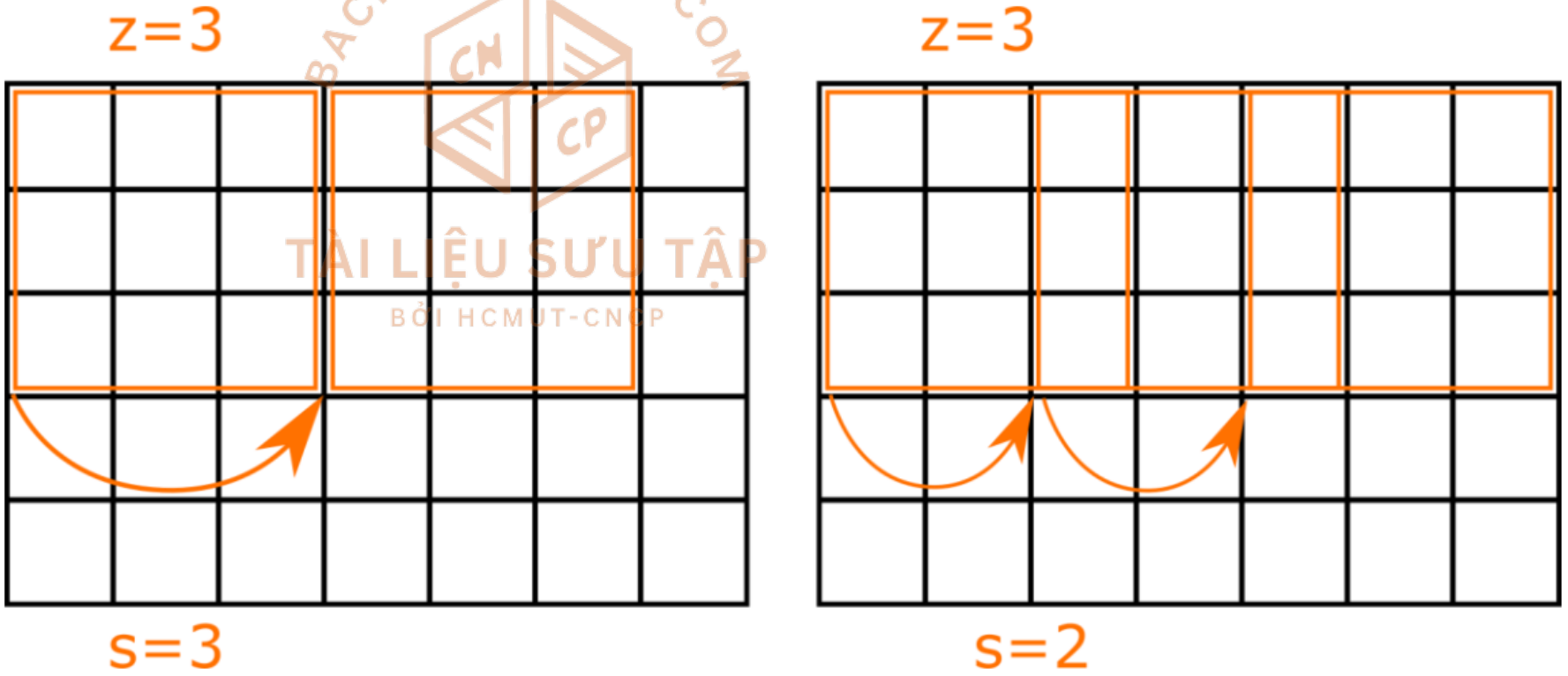


Architecture

- ✓ Training on Multiple GPUs
  - ❖ Smaller error

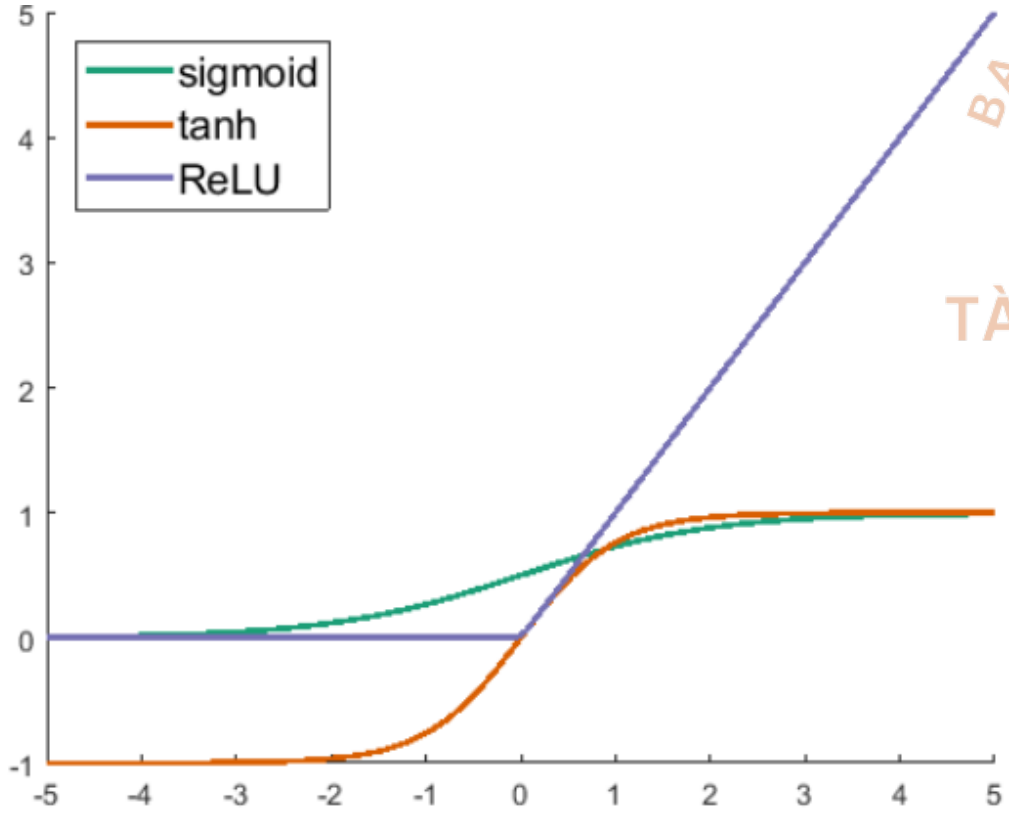
Architecture

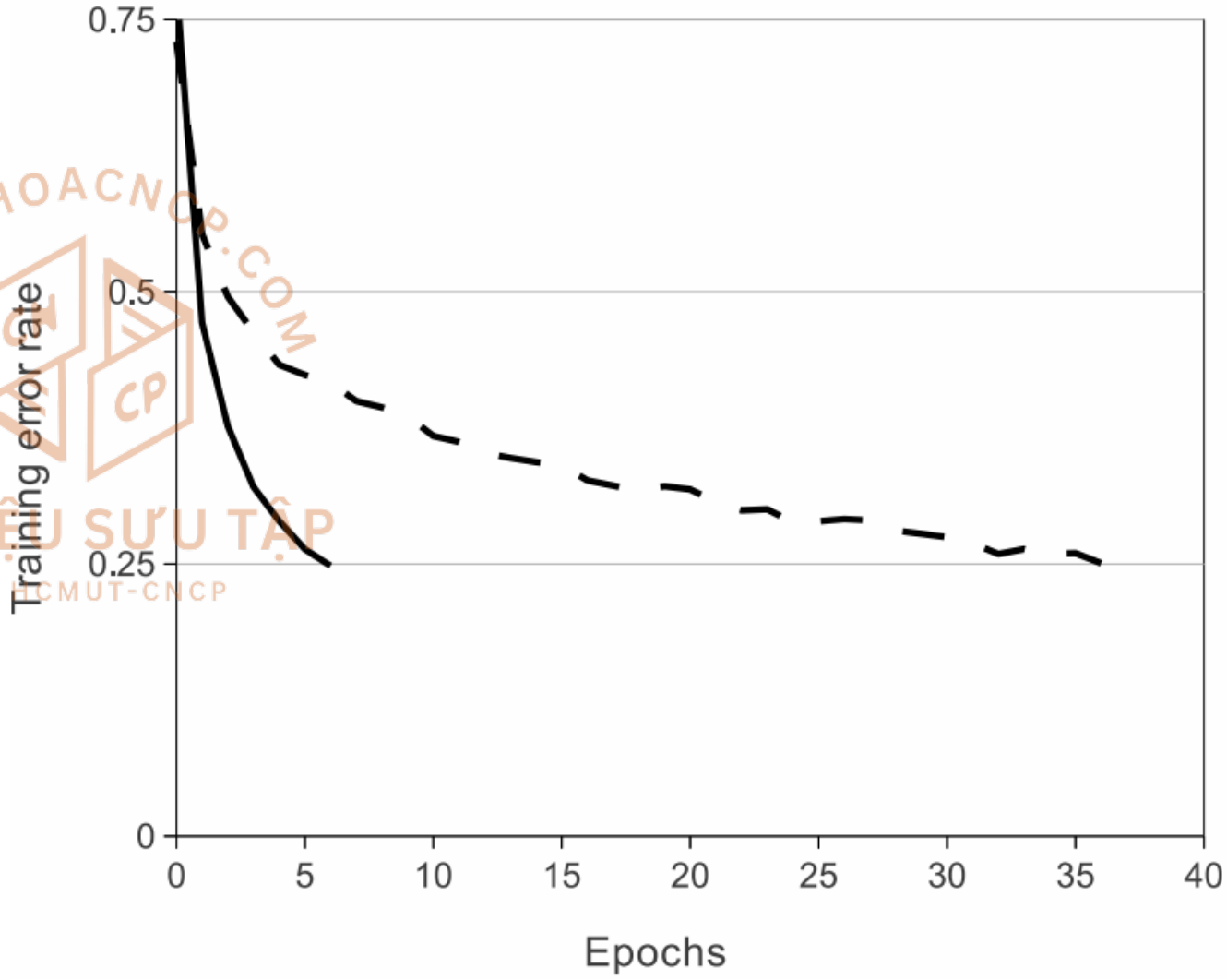
- Pooling
  - Overlapping pooling
  - Window: 3x3
  - Stride: 2



Architecture

- ✓ Activation function
  - ReLU
  - ReLU vs tanh





Overfitting

- Data augmentation
  - Without data augmentation: substantial overfitting
  - Image translation & horizontal reflection
    - 256x256 image → 224x224 patches
    - 2048 times
  - Altering intensities of RGB channels
- Test time: 1 images → 5 patches → 10 patches (reflection)

Overfitting

- Dropout
  - Without dropout: substantial overfitting
  - Used in the first two fully-connected layers
  - $p = 0.5$
  - Roughly doubles # iterations

Training

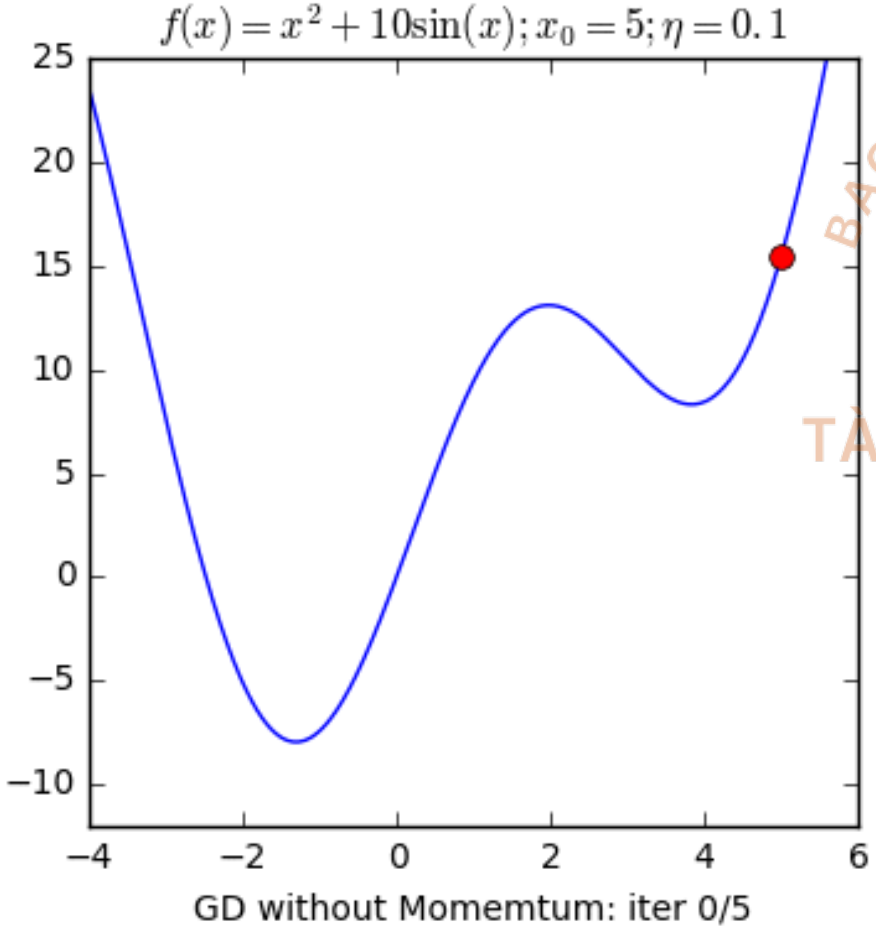
- ✓ Stochastic gradient descent
- ✓ Actually: mini-batch gradient descent
- ✓ With momentum
- ✓ Batch size: 128
  - ❖ Momentum: 0.9
  - ❖ Weight decay: 0.0005 (for regularization?)

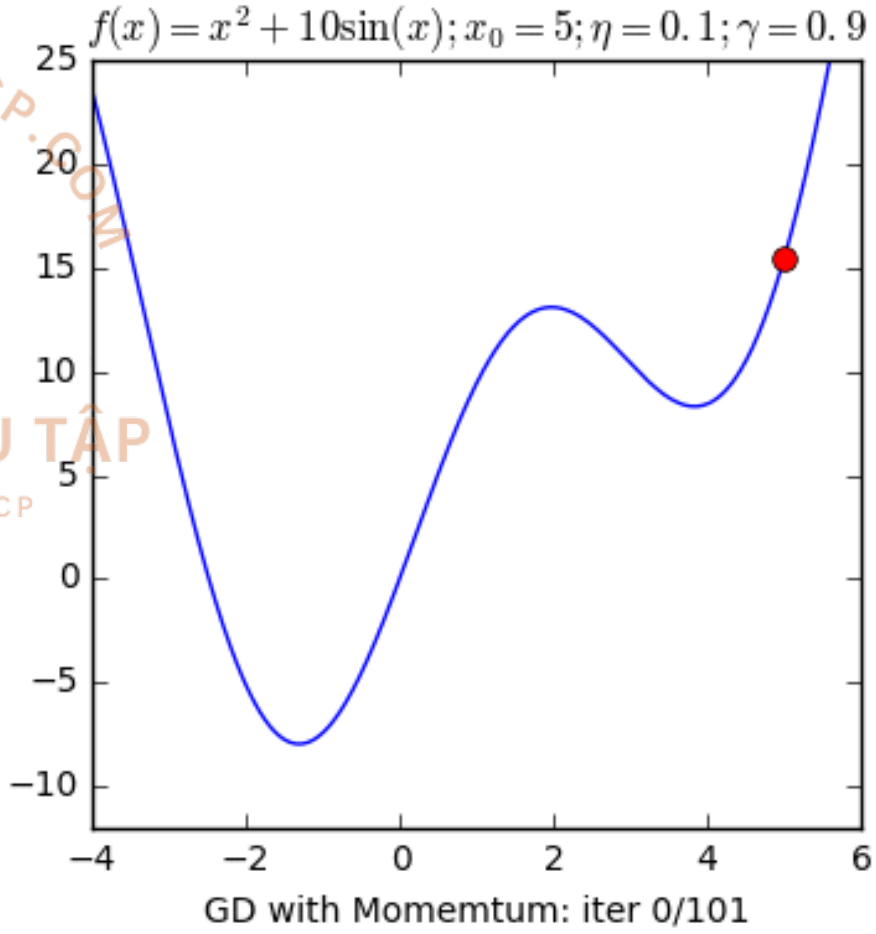
$$v_{i+1} := 0.9 \cdot v_i - 0.0005 \cdot \epsilon \cdot w_i - \epsilon \cdot \left\langle \frac{\partial L}{\partial w}\Big|_{w_i} \right\rangle_{D_i}$$

$$w_{i+1} := w_i + v_{i+1}$$

Results

- Example: $f(x) = x^2 + 10\sin(x)$





Training

We initialized the weights in each layer from a zero-mean Gaussian distribution with standard deviation 0.01. We initialized the neuron biases in the second, fourth, and fifth convolutional layers, as well as in the fully-connected hidden layers, with the constant 1. This initialization accelerates the early stages of learning by providing the ReLUs with positive inputs. We initialized the neuron biases in the remaining layers with the constant 0.

We used an equal learning rate for all layers, which we adjusted manually throughout training. The heuristic which we followed was to divide the learning rate by 10 when the validation error rate stopped improving with the current learning rate. The learning rate was initialized at 0.01 and reduced three times prior to termination. We trained the network for roughly 90 cycles through the training set of 1.2 million images, which took five to six days on two NVIDIA GTX 580 3GB GPUs.

Results

- ✓ Evaluation criteria
  - ❖ Top-1 & top-5 error rate
- ✓ ILSVRC-2010

Model	Top-1	Top-5
Sparse coding [2]	*47.1%*	*28.2%*
SIFT + FVs [24]	*45.7%*	*25.7%*
CNN	**37.5%**	**17.0%**

- ✓ ILSVRC-2012

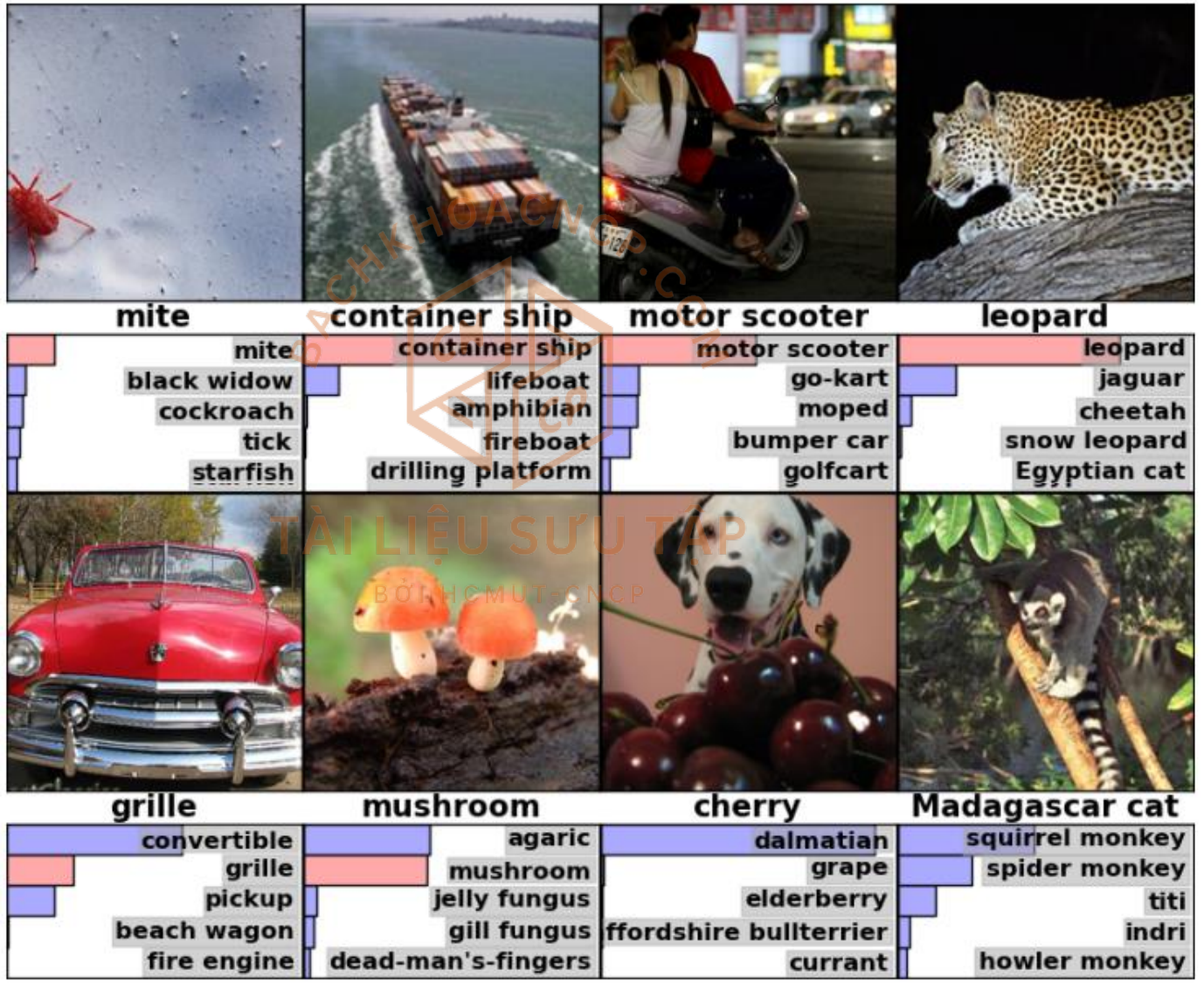
Model	Top-1 (val)	Top-5 (val)	Top-5 (test)
SIFT + FVs [7]	—	—	*26.2%*
1 CNN	40.7%	18.2%	—
5 CNNs	38.1%	16.4%	**16.4%**
1 CNN*	39.0%	16.6%	—
7 CNNs*	36.7%	15.4%	**15.3%**

Results

- ImageNet 2009
  - 10.184 classes
  - 8.9M images
  - 6th convolutional layer added

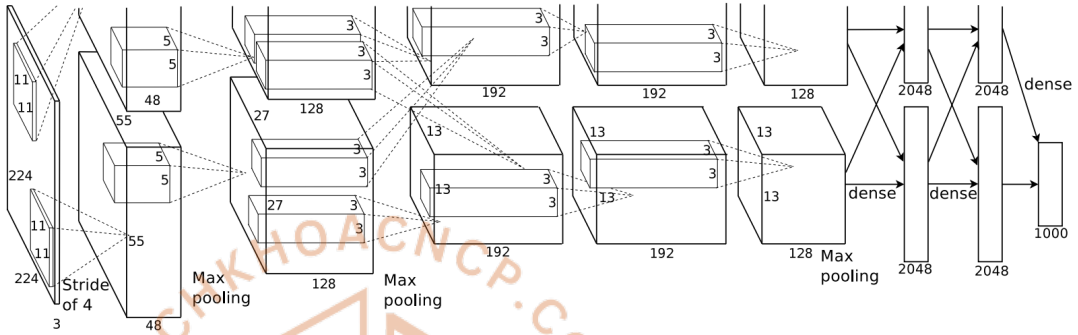
Model	Top 1 error rate (%)	Top 5 error rate (%)
AlexNet	67.4	40.9
Other	78.1	60.9

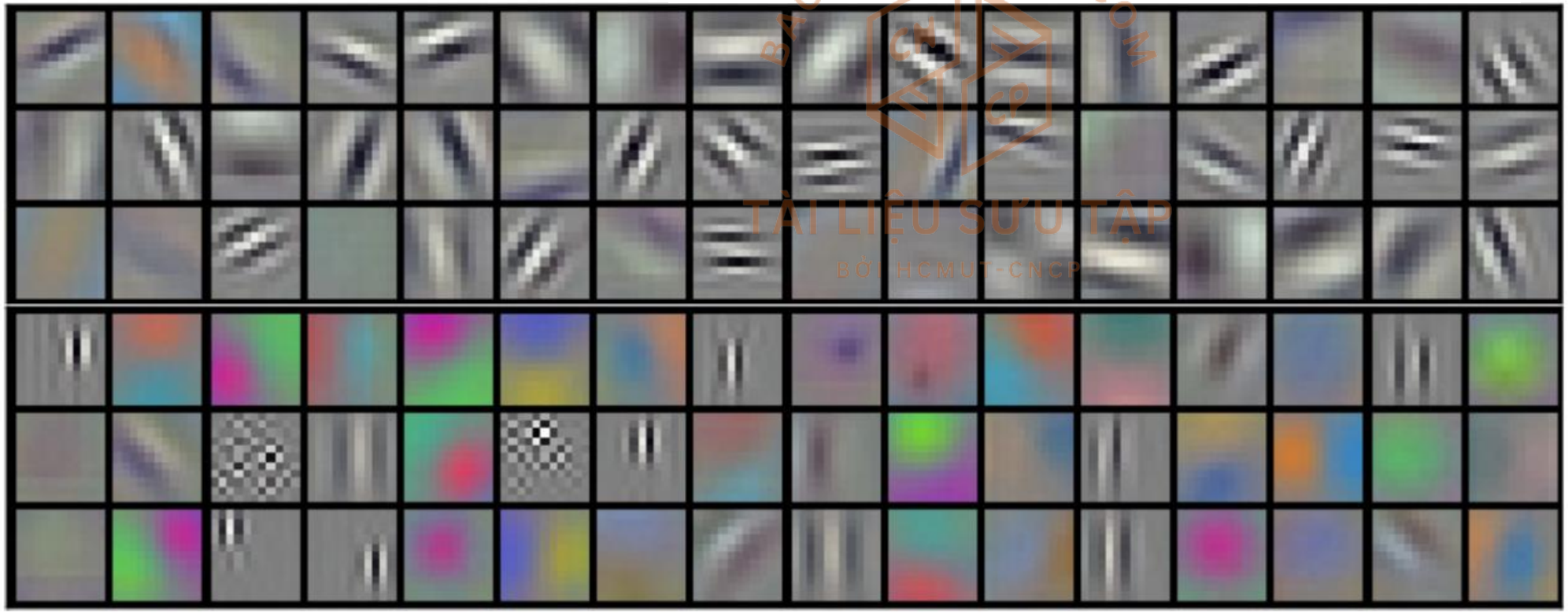
Results



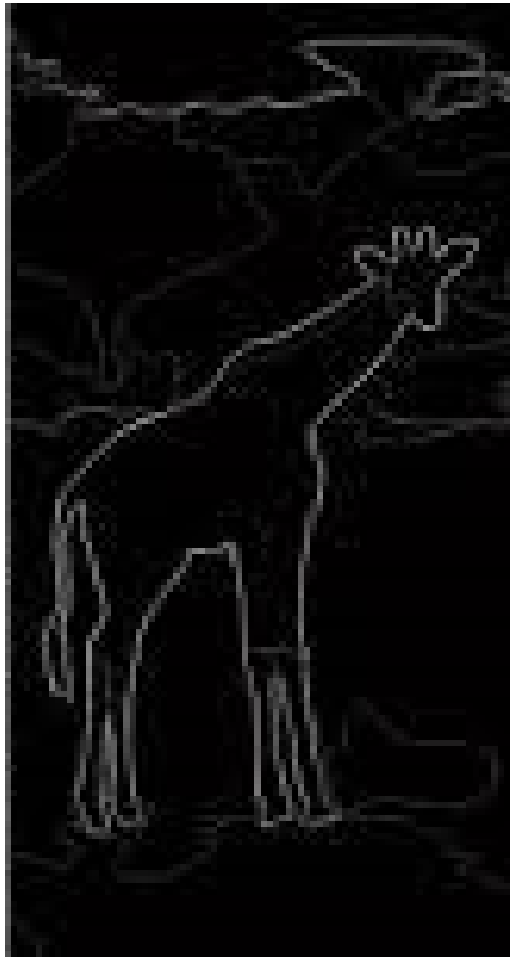
Results

- ✓ Kernel: 96, 11x11x3





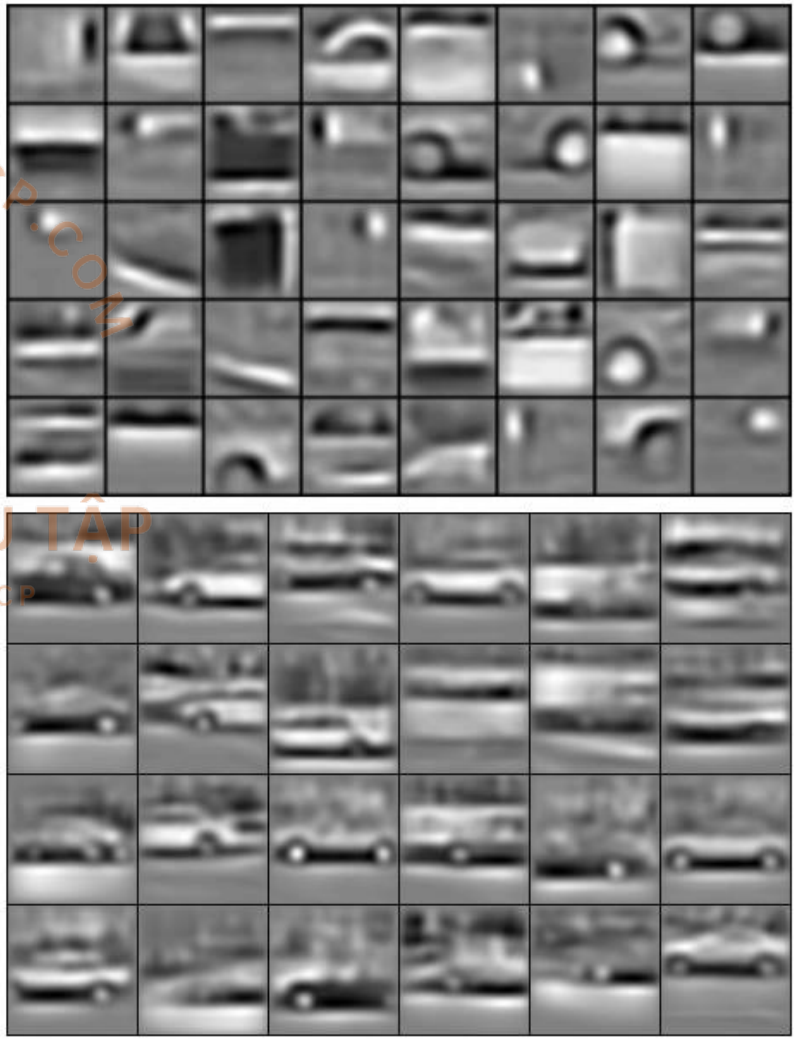
Results



Results

faces

cars

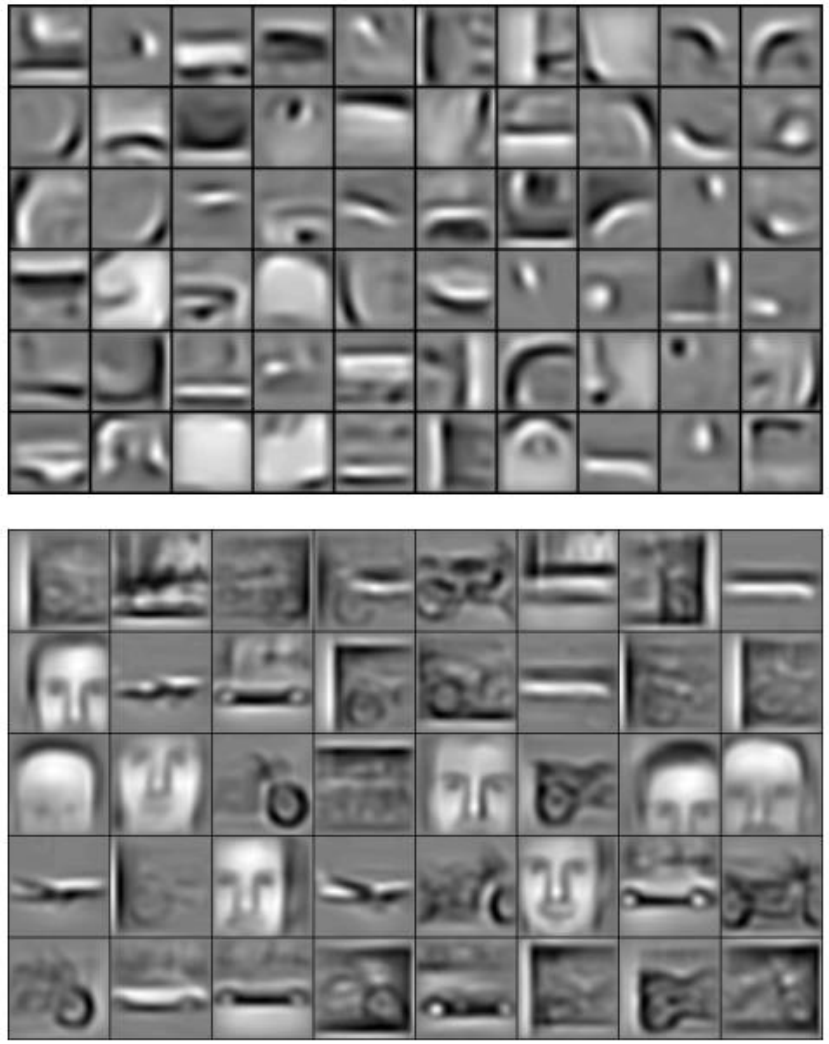


Results

elephants

chairs

faces, cars, airplanes, motorbikes



Results